(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 489 596 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
22.12.2004 Bulletin 2004/52

(51) Int Cl.[7]: **G10L 11/02**, H04R 3/00, H04R 1/40

(21) Application number: **03445076.7**

(22) Date of filing: **17.06.2003**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PT RO SE SI SK TR
Designated Extension States:
AL LT LV MK

(71) Applicant: **Sony Ericsson Mobile Communications AB**
221 88 Lund (SE)

(72) Inventor: **Gustavsson, Stefan**
254 43 Helsingborg (SE)

(74) Representative: **Akerman, Marten Lennart et al**
Albihns Malmö AB
Box 4289
203 14 Malmö (SE)

(54) Device and method for voice activity detection

(57) The invention relates to a device, a mobile apparatus incorporating the device, an accessory therefore and a method for voice activity detection, particularly in a mobile telephone, using the directional sensitivity of a microphone system and exploiting the knowledge about the voice source's orientation in space. The device comprises a sound signal analyser arranged to determine whether a sound signal comprises speech. According to the invention, the device further comprises a microphone system (2a,2b,2c,2d,2e) arranged to discriminate sounds emanating from sources located in different directions from the microphone system, so that sounds only emanating from a range of directions are included as signals possibly containing speech.

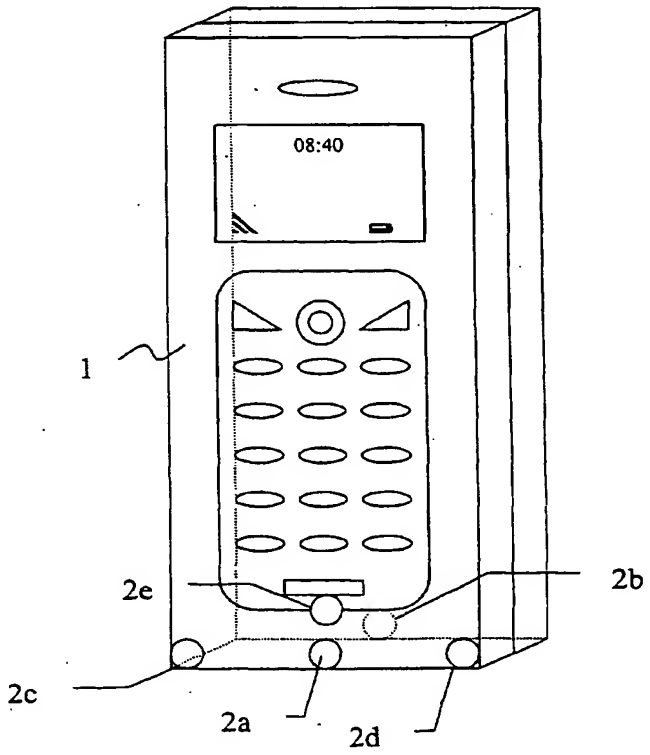

FIG 1

EP 1 489 596 A1

Description

Field of the invention

[0001] The present invention relates to a device, a mobile apparatus incorporating the device, an accessory therefor and a method for voice activity detection, particularly in a mobile telephone, using the directional sensitivity of a microphone system and exploiting the knowledge about the voice source's orientation in space. The device assists the existing voice activity detection to achieve higher sensitivity and requiring less processor power.

State of the art

[0002] Voice activity detectors are used e.g. in mobile phones to enhance the performance in certain situations. The most common way to construct a voice activity detector is to look at the levels of the sub-bands of the incoming signal. Then the background noise level and the speech level are estimated and compared with a threshold to determine whether speech is present or not. An example of a voice activity detector is disclosed in U.S. patent 6,427,134.

[0003] For instance in noisy environments it is hard to make a uniform parameter set-up for the voice activity detector. Therefore several voice activity detectors are needed, trimmed to the specific cases. For example in some modules you need to be sure that if there is speech it should be detected (echo canceller), but in other cases it is better to indicate no speech if the signal to noise ratio level is too low. The plurality of voice activity detectors put a load on the digital signal processors that have to take care of performing the various voice activity detection algorithms.

Summary of the invention

[0004] An object of the present invention is to complement existing voice activity detection taking into account the direction of the source of the sound.

[0005] In a first aspect, the invention provides a device for voice activity detection comprising a sound signal analyser arranged to determine whether a sound signal comprises speech.

[0006] According to the invention, the device further comprises

a microphone system arranged to discriminate sounds emanating from sources located in different directions from the microphone system, so that sounds only emanating from a range of directions are included as signals possibly containing speech.

[0007] Suitably, the range of directions is directed in the direction of an intended user's mouth.

[0008] In one embodiment, the microphone system comprises two microphone elements separated a distance and located on a line directed in the direction of an intended user's mouth.

[0009] The range of directions may be defined as all sounds falling inside a cone with a cone angle α, wherein $10° < \alpha < 30°$, and preferably, α is approximately 25°.

[0010] In another embodiment, the microphone system comprises three microphone elements separated a distance and located in a plane directed in the direction of an intended user's mouth.

[0011] Suitably, two of said three microphone elements are separated a distance and located on a line directed perpendicular to the direction of an intended user's mouth.

[0012] In another embodiment, the microphone system comprises four microphone elements located such that the fourth microphone is not located in the same plane as the three others.

[0013] The microphone elements may be directional with a pattern having maximal sensitivity in the direction of an intended user's mouth.

[0014] In still a further embodiment, the microphone system comprises one directional microphone element together with one or more other microphone elements to remove the uncertainty in the direction of the sound source. The directional microphone element may be used to measure the sound pressure level relative to the other microphone element.

[0015] In a second aspect, the invention provides a mobile apparatus comprising a device as mentioned above.

[0016] Suitably, the microphone elements are located at the lower edge of the apparatus.

[0017] In one embodiment, a plurality of microphone elements are located at the lower edge of the apparatus and at least one further microphone element is located at a distance from the lower edge.

[0018] The mobile apparatus may be a mobile radio terminal, e.g. a mobile telephone, a pager, a communicator, an electric organiser or a smartphone.

[0019] In a third aspect, the invention provides an accessory for a mobile apparatus comprising a microphone system as mentioned above.

[0020] Suitably, the direction of the range of directions is adjustable.

[0021] The accessory may be a hands-free kit or a telephone conference microphone.

[0022] In a fourth aspect, the invention provides a method for voice activity detection, including the steps of:

receiving sound signals from a microphone system arranged to discriminate sounds emanating from sources located in different directions from the microphone system;

determining the direction of the sound source causing the sound signals;

if the sounds emanate from a first range of directions, further analyse

the sound to determine whether the sound signal

comprises speech;
but if the sounds emanate from a second, different range of directions decide that the sound signal does not comprise speech.

[0023] Suitably, the first range of directions is directed in the direction of an intended user's mouth.

[0024] The first range of directions may be defined as all sounds falling inside cone with a cone angle α, wherein $10° < \alpha < 30°$, and preferably α is approximately 25°.

[0025] In one embodiment, the microphone system comprises at least two microphone elements located at a distance from each other and located on a line directed in the direction of an intended user's mouth, said two microphone elements being separated a distance d, wherein the direction to the sound source θ is calculated as

$$\theta = \arccos \frac{\Delta t \cdot v}{2 \cdot d}$$

where
Δt is the time difference between the sounds from the two microphone elements,
v is the velocity of sound.

[0026] In another embodiment, one directional microphone element is used together with one or more other microphone elements to remove the uncertainty in the direction of the sound source.

[0027] The directional microphone element may be used to measure the sound pressure level relative to the other microphone element.

[0028] The invention is defined in the attached independent claims 1, 12, 16, and 20, while preferred embodiments are set forth in the dependent claims.

Brief description of the drawings

[0029] The invention will be described below in greater detail with reference to the accompanying drawings, in which:

fig. 1 is a perspective view of a mobile phone incorporating the present invention, and
fig. 2 is a schematic drawing of the receiving angle of an embodiment of the present invention.

Detailed description of preferred embodiments

[0030] As mentioned briefly in the introduction, many signal processing algorithms, such as echo cancellation and background noise synthesis, used in phones and hands-free kits are based on the fact that the user is speaking or not. For example the speech codec is active when the near-end user is speaking and the background synthesis is active when the near-end user is silent. All these algorithms need good voice activity detectors (VAD) to perform well. An error in the detection can result in artefacts or malfunctions caused by divergence of the algorithms or other problems.

[0031] Existing voice activity detectors are directed to determine whether speech is present or not in a sound signal. However, in fact not all speech is interesting or relevant, but only the user's speech. All other speech, e.g. in a noisy environment with several persons speaking, could be ignored and regarded as just noise.

[0032] The present inventor has realised that a microphone system having some kind of directional sensitivity could be used to discriminate sound emanating from different sources located in different directions. Sound not emanating from the user can be declared as non-speech, and those signals do not have to be analysed with the conventional voice activity detectors.

[0033] The existing voice activity detectors may be conventional and are only referred to as a sound signal analyser in this application.

[0034] Generally, a microphone system having some kind of directional sensitivity can be used. Fig. 1 shows an example with at least two separate microphone elements.

[0035] A general mobile telephone is indicated at 1. The invention is equally applicable to other devices such as mobile radio terminals, pagers, communicators, electric organisers or smartphones. The common feature is that voice activity detection is employed, e.g. in connection with communicating speech or receiving voice commands by means of speech recognition.

[0036] In the simplest version, the microphone system comprises two microphones 2a and 2b. Suitably, they are located on a line directed in the calculated direction of an intended user's mouth. Suitably, the microphone elements are located at the lower edge of the mobile apparatus 1.

[0037] Fig. 2 shows a schematic diagram of the calculation of the direction of the sound source, typically the user's mouth 3. In the case of two microphones, only the angle to the line on which the microphone elements are located can be determined. In other words, the direction of the sound source is on a cone with a cone angle θ. To calculate the angle θ, first a cross-correlation between the two signals from the microphones 2a and 2b is made. The maximum indicates the time difference Δt between the two microphones 2a and 2b. The distance between the two microphones 2a and 2b is e.g. 20 millimetres. The angle θ is calculated as

$$\theta = \arccos \frac{\Delta t \cdot v}{2 \cdot d}$$

[0038] Note that arccos is only defined for arguments between —1 and 1. If the time difference is negative, this means that the angle is greater than 90° and the sound emanates from behind the apparatus.

[0039] Suitably, the device is adapted to determine that all sounds with an angle θ less than a fixed angle α

are emanating from the user. The threshold angle α may be set within a range of e.g. 10° to 30°, suitably at 25°.

[0040] In the case of three microphones, the direction of the sound source can be further determined to be at two points (e.g. on the above cone). The three microphone elements are suitably located in a plane directed in the general direction of the user's mouth. In fig. 1 microphone elements 2b, 2c and 2d are a possible set-up. The two microphone elements 2c and 2d at the front are located on a line perpendicular to the direction of the user's mouth, while the third microphone element 2b is located at the rear side.

[0041] In the case of four microphones (or more) detection of all direction angles may be calculated, provided that four microphone elements are located such that the fourth microphone is not located in the same plane as the three others, e.g. on a tetrahedron. A possible set-up is two microphone elements 2c and 2d at the front on the lower edge, while a third microphone element 2b is located at the rear side, and a fourth microphone element 2e is located at the front at a distance from the lower edge.

[0042] A similar microphone arrangement may be used in an accessory to a mobile apparatus, such as a hands-free kit or a telephone conference microphone system intended to be placed on a table. Apart from the microphone elements the logic circuitry may be located in the main/mobile apparatus. In this case the reception angle of the microphone system can be adjustable. This is useful e.g. when the microphone system is placed in a car, where the user can be seated either in the driver's seat or in the passenger's seat or even both the driver and the passenger may be speakers during the same call. The adjustment of the reception angle can be achieved mechanically or electronically, for example by beam forming or adaptation of the directional sensitivity of the microphone system.

[0043] To further enhance the sensitivity of the microphone system, directional microphone elements with a pattern having a maximum sensitivity in the direction of the user's mouth could be used.

[0044] In a further embodiment, one directional microphone element is used together with one or two other microphone elements (that may be non-directional). The directional microphone element is used to measure the sound pressure level relative to the other(s), thus removing the uncertainty in the direction of the sound source. Various combinations of directional microphone elements and non-directional microphone elements are possible.

[0045] The present invention leads to a voice activity detector having enhanced performance. With the present invention only one voice activity detector may be necessary throughout the whole signal path. This will in turn reduce the computational complexity, decreasing the load on the digital signal processors as well as improving the performance. It is especially favourable in environments with high background noise and noise with similar spectral properties as speech.

[0046] A person skilled in the art will realise that the invention may be realised with various combinations of hardware and software. The scope of the invention is only limited by the claims below.

Claims

1. A device for voice activity detection comprising a sound signal analyser arranged to determine whether a sound signal comprises speech, **characterised by**
   a microphone system (2a, 2b, 2c, 2d, 2e) arranged to discriminate sounds emanating from sources located in different directions from the microphone system, so that sounds only emanating from a range of directions are included as signals possibly containing speech.

2. A device according to claim 1, **characterised in that** the range of directions is directed in the direction of an intended user's mouth (3).

3. A device according to claim 2, **characterised in that** the microphone system comprises two microphone elements (2a, 2b) separated a distance and located on a line directed in the direction of an intended user's mouth (3).

4. A device according to claim 3, **characterised in that** the range of directions is defined as all sounds falling inside a cone with a cone angle α, wherein $10° < \alpha < 30°$.

5. A device according to claim 3, **characterised in that** α is approximately 25°.

6. A device according to claim 2, **characterised in that** the microphone system comprises three microphone elements (2b, 2c, 2d) separated a distance and located in a plane directed in the direction of an intended user's mouth (3).

7. A device according to claim 6, **characterised in that** two (2c, 2d) of said three microphone elements are separated a distance and located on a line directed perpendicular to the direction of an intended user's mouth (3).

8. A device according to claim 2, **characterised in that** the microphone system comprises four microphone elements (2b, 2c, 2d, 2e), located such that the fourth microphone (2e) is not located in the same plane as the three others (2b, 2c, 2d).

9. A device according to any one of claims 1 to 8, **characterised in that** the microphone elements (2a, 2b,

2c, 2d, 2e) are directional with a pattern having maximal sensitivity in the direction of an intended user's mouth (3).

10. A device according to claim 1, **characterised in that** the microphone system comprises one directional microphone element together with one or more other microphone elements adapted to remove the uncertainty in the direction of the sound source.

11. A device according to claims 10, **characterised in that** the directional microphone element is adapted to measure the sound pressure level relative to the other microphone element.

12. A mobile apparatus, **characterised in that** it comprises a device as defined in any one of claims 1 to 11.

13. A mobile apparatus according to claim 12, **characterised in that** the microphone elements (2a, 2b, 2c, 2d) are located at the lower edge of the apparatus.

14. A mobile apparatus according to claim 12, **characterised in that** a plurality of microphone elements (2a, 2b, 2c, 2d) are located at the lower edge of the apparatus and at least one further microphone element (2e) is located at a distance from the lower edge.

15. A mobile apparatus according to any one of claims 12 to 14, **characterised in that** it is a mobile radio terminal, e.g. a mobile telephone (1), a pager, a communicator, an electric organiser or a smartphone.

16. An accessory for a mobile apparatus, **characterised in that** it comprises a microphone system (2a, 2b, 2c, 2d, 2e) as defined in any one of claims 1 to 11.

17. An accessory according to claim 16, **characterised in that** the direction of the range of directions is adjustable.

18. An accessory according to claim 16 or 17, **characterised in that** it is a hands-free kit.

19. An accessory according to claim 16 or 17, **characterised in that** it is a telephone conference microphone.

20. A method for voice activity detection, **characterised by** the steps of:

receiving sound signals from a microphone system (2a, 2b, 2c, 2d, 2e) arranged to discriminate sounds emanating from sources located in different directions from the microphone system;
determining the direction of the sound source causing the sound signals;
if the sounds emanate from a first range of directions, further analyse the sound to determine whether the sound signal comprises speech;
but if the sounds emanate from a second, different range of directions decide that the sound signal does not comprise speech.

21. A method according to claim 20, **characterised in that** the first range of directions is directed in the direction of an intended user's mouth (3).

22. A method according to claims 21, **characterised in that** the first range of directions is defined as all sounds falling inside a cone with a cone angle α, wherein $10° < \alpha < 30°$.

23. A method according to claims 22, **characterised in that** α is approximately 25°.

24. A method according to any one of claims 22 or 23, **characterised in that** the microphone system comprises at least two microphone elements (2a, 2b) located at a distance from each other and located on a line directed in the direction of an intended user's mouth (3), said two microphone elements being separated a distance d, wherein the direction to the sound source θ is calculated as

$$\theta = \arccos \frac{\Delta t \cdot v}{2 \cdot d}$$

where
Δt is the time difference between the sounds from the two microphone elements,
v is the velocity of sound.

25. A method according to claims 20, **characterised in that** one directional microphone element is used together with one or more other microphone elements to remove the uncertainty in the direction of the sound source.

26. A method according to claims 25, **characterised in that** the directional microphone element is used to measure the sound pressure level relative to the other microphone element.

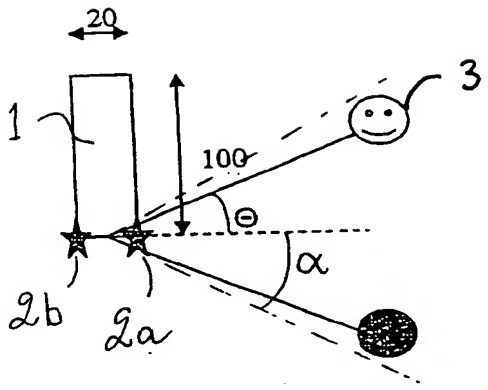

FIG 2

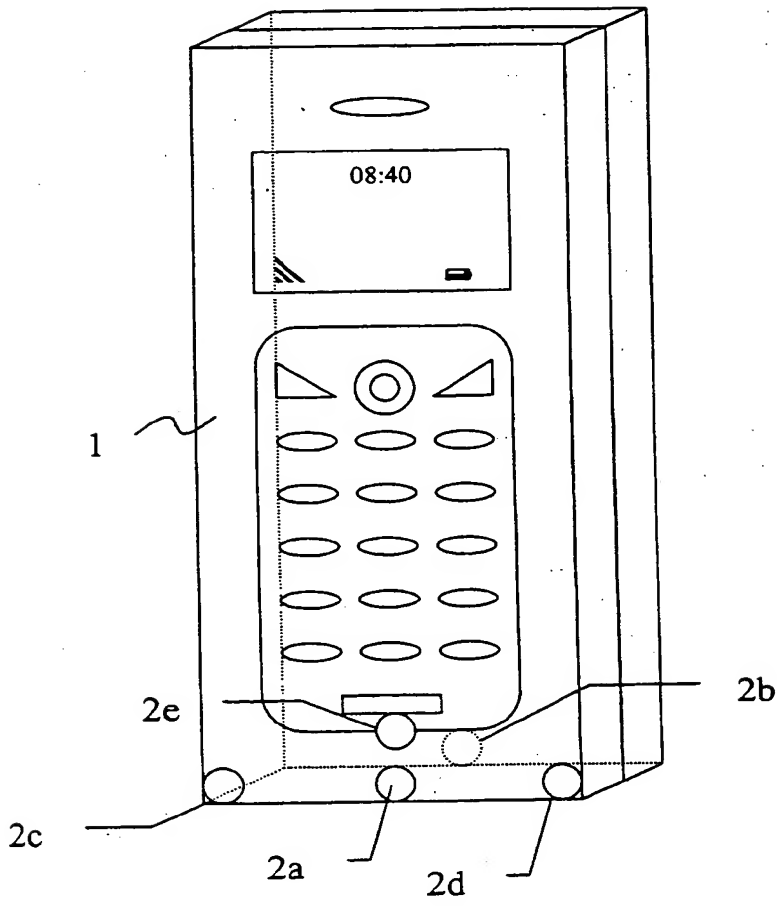

FIG 1

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 03 44 5076

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	US 2003/027600 A1 (ORAINTARA SOOTORN ET AL) 6 February 2003 (2003-02-06) * paragraph [0006] * * paragraph [0017] * * paragraph [0022] *	1,2,9, 11,12, 20,21,25	G10L11/02 H04R3/00 H04R1/40
Y	idem	3,6-8, 16-19	
Y	US 2002/009203 A1 (ERTEN GAMZE) 24 January 2002 (2002-01-24) * figures 2,4 * * paragraph [0009] * * paragraphs [0044],[0045] *	3	
Y	* figures 9,12 *	16-19	
Y	EP 1 206 161 A (SONY INTERNAT EUROP GMBH) 15 May 2002 (2002-05-15) * abstract * *, sentence 40 - sentence 44 * * paragraph [0021] * * paragraphs [0027],[0028] * * figures 2,3 *	6-8	

TECHNICAL FIELDS SEARCHED (Int.Cl.7)
G10L
H04R

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
MUNICH	10 October 2003	Krembel, L

CATEGORY OF CITED DOCUMENTS
X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 03 44 5076

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

10-10-2003

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2003027600 A1	06-02-2003	NONE	
US 2002009203 A1	24-01-2002	AU 5121301 A	15-10-2001
		CA 2404071 A1	11-10-2001
		CN 1436436 T	13-08-2003
		EP 1295507 A2	26-03-2003
		WO 0176319 A2	11-10-2001
EP 1206161 A	15-05-2002	EP 1206161 A1	15-05-2002

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82